PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/11, C12P 21/08
C12Q 1/68, C12N 15/62
C07K 15/28

A2

(11) International Publication Number: WO 94/01548

(43) International Publication Date: 20 January 1994 (20.01.94)

(21) International Application Number: PCT/GB93/01467

(22) International Filing Date: 13 July 1993 (13.07.93)

(30) Priority data:
9214857.6 13 July 1992 (13.07.92) GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : SIBSON, David, Ross [GB/GB]; 37 Grimsdells Lane, Amersham, Buckinghamshire HP6 6HF (GB). GROSS, Jacqueline [GB/GB]; 47 Boxmoor Road, Kenton, Middlesex HA3 8LH (GB). HADFIELD, Kathryn, Mary [GB/GB]; 5 Carlisle Terrace, St Ives, Huntingdon, Cambridgeshire PE17 4PQ (GB). HOWELLS, David [GB/GB]; 77 Puttocks Drive, Welham Green, Hatfield, Hertfordshire AL9 7LW (GB). STARKEY, Michael [GB/GB]; 27 Creasy Close, Abbots Langley, Hertfordshire WO5 0HS (GB). KELLY, Maria [IE/GB]; 24A Oxford Road, Ealing, London W5 3ST (GB). SHAW, Diana [GB/CA]; 342 Glacier Hall, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4 (CA).

(74) Agent: BIZLEY, Richard, Edward; Hepworth Lawrence Bryer & Bizley, 2nd Floor Gate House South, West Gate, Harlow, Essex CM20 1JN (GB).

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE, PLACENTA OR BONE NARROW

(57) Abstract

This invention provides a nucleic acid fragment encoding a gene product or portion thereof and comprising any one of: (a) a sequence selected from SEQ ID Nos 1 to 1193 from the attached sequence listings; (b) an allelic variation of a sequence as defined in (a); or (c) a sequence complementary to (a) or (b). The invention includes uses of such fragments, and gene products corresponding thereto.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NE	Niger
BE	Belgium	GN	Guinea	NL	Netherlands
BF	Burkina Faso	GR	Greece	NO	Norway
BG	Bulgaria	HU	Hungary	NZ	New Zealand
BJ	Benin	IE	Ireland	PL	Poland
BR	Brazil	IT	Italy	PT	Portugal
BY	Belarus	JP	Japan	RO	Romania
CA	Canada	KP	Democratic People's Republic of Korea	RU	Russian Federation
CF	Central African Republic			SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovak Republic
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	UA	Ukraine
DE	Germany	MG	Madagascar	US	United States of America
DK	Denmark	ML	Mali	UZ	Uzbekistan
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE, PLACENTA OR BONE NARROW

This invention relates to new nucleic acid fragments encoding gene products or portions thereof, which fragments are obtainable from human nucleic acid populations, individual members of such populations being present in widely varying amounts.

Situations are increasingly arising in which it is necessary to study complex nucleic acid or polynucleotide populations. For example, it is now widely appreciated that an invaluable resource could be created if the entire sequence of the genomes of organisms such as man were determined and the information available. The magnitude of such a task should not, however, be underestimated. Thus, the human genome may contain as many as 100,000 genes [a very substantial proportion of which may be expressed in the human brain (Sutcliffe, Ann. Rev. Neurosci. 11:157 (1988))]. Only a very small percentage of the stock of human genes has presently been explored, and this largely in a piecemeal and usually specifically targeted fashion.

There has been much public debate about the best means of approaching human genome sequencing. Brenner has argued (CIBA Foundation Symposium 149:6 (1990)) that efforts should be concentrated on cDNAs produced from reverse transcribed mRNAs rather than on genomic DNA. This is primarily because most useful genetic information resides in the fraction of the genome which corresponds to mRNA, and this fraction is a very small part of the total (5% or less). Moreover, techniques for generating cDNAs are also well known. On the other hand, even supposing near perfect recovery of cDNAs corresponding to all expressed mRNAs, some potentially useful information will be lost by the cDNA approach, including sequences responsible for control and regulation of genes. Nonetheless, the cDNA approach at least substantially reduces the inherent inefficiencies resulting from analysis of repeated sequences or non-coding sequences in an approach which depends upon genomic DNA sequencing.

Recently, the results of a rapid method for identifying and characterising new cDNAs has been reported (Adams, M.D. et al., Science 252, 1991, pp 1651-1656). Essentially, a semi-automated sequence reader was used to produce a single read of sequence from one end of each of a number of cDNAs picked at random. It was shown, by comparing the nucleic acid sequences of the cDNAs (or the protein sequences produced by translating the nucleic acid sequence of the cDNAs) to each

other and to known sequences in public databases, that each of the cDNAs picked at random, could be unambiguously classified. The cDNAs could be classified as being either entirely new or as corresponding, to a greater or lesser extent, to a previously known sequence. cDNAs identified in this way were further characterised and found to be useful in a variety of standard applications, including physical mapping. Unfortunately, such a process is insufficient. The longer the process is pursued with any given population of cDNAs the less efficient it becomes and the lower the rate of identification of new clones. In essence, as the number of cDNAs which have already been picked rises, the probability of picking a particular cDNA more than once increases. This difficulty is exacerbated by the wide range of abundancies at which different cDNAs can occur, which abundancies can vary by several orders of magnitude. Thus, whereas some sequences are exceedingly rare, a single cDNA type may comprise as much as 10% of the population of cDNAs produced from a particular tissue (Lewin, B. Gene Expression, Vol. 2: Eukaryotic Chromosomes, 2nd ed., pp. 708-719. New York: Wiley, 1980). The need to avoid missing rarer species in any given population presents a considerable problem.

Various approaches (so-called "normalisation" techniques) have been tried in addressing the problem of increasing the efficiency of examination of a mixed nucleotide population, for example, such a population as is to be examined in human genome sequencing.

Thus, a standard PCR protocol can be used to amplify selectively cDNAs which are present at extremely low levels, if there is information about the sequence of those cDNAs. If not, a primer specific to the desired cDNA cannot be constructed and the desired cDNA cannot be selectively amplified. The standard PCR method is therefore inadequate if it is desired to characterise a number of unknown genes.

A second approach involves hybridization of cDNA to genomic DNA. At saturation, the cDNAs recovered from genomic/cDNA hybrids will be present in the same abundance as the genes encoding them. This will provide a much more homogenous population than the original cDNA library, but does not entirely solve the problem. In order to reach saturation in respect of the very rare sequences, it will be necessary to use huge quantities of cDNA, which need to be allowed to anneal to large amounts of genomic DNA over a considerable periods of time. Furthermore, cDNAs which are homologous to genes which are present in multiple copies in the genome will be over-represented.

A third approach exploits the second order reassociation kinetics of cDNA annealing to itself. After a long period of annealing, the cDNAs which remain single stranded will have nearly the same abundance, and can be recovered by standard PCR (see Patanjali, S.R. et al., PNAS USA 88, 1991, pp. 1943-1947; Ko, M.S.H., NAR 19, no.18, 1991, pp 5705-5711). The methods disclosed in these two publications, however, suffer from notable disadvantages. They are entirely dependent on the stringent physical separation of single stranded and double stranded DNA, require an elevated number of manual manipulations in each reaction, and necessitate protracted hybridisation times (up to 288 hours in the method of Patanjali et al.)

Yet a further approach in "normalising" a nucleotide population is described in co-pending British Patent Application No 91 15407.0, filed 17th July, 1991 by MRC, and involves a PCR process in which a mixture comprising a heterogenous DNA population and appropriate oligonucleotide primers is first formed and the DNA denatured, but before effecting a conventional PCR protocol the conditions are altered to allow the denatured strands of the more common DNA species to reanneal together, whilst avoiding annealing of primers to the DNA strands. By this means, rarer species can subsequently be amplified in preference to the more common species.

This PCR normalisation method in general comprises the steps of:

(a) preparing a mixture comprising a heterogenous DNA population and oligonucleotide primers suitable for use in a PCR process, in which the DNA is denatured;

(b) altering the conditions to allow the denatured strands of the more common DNA species to reanneal, while preventing the annealing to the primers to the DNA strands;

(c) further altering the conditions of the mixture in order to allow the primers to anneal to the remaining single-stranded DNA comprising the rarer DNA species; and

(d) carrying out an extension synthesis in the mixture produced in step (c).

Advantageously, the method consists of a cyclic application of the above four steps.

It will be appreciated that the conditions may be altered by the alteration of the temperature of the reaction mixture. However, any conditions which affect the hybridisation of complementary DNA strands to one another may be varied to achieve the required result.

Because the reannealing efficiency of any given DNA species will depend on the product of its concentration and time, the more abundant the sequence the greater the extent to which it will reanneal in any given time period. Once a DNA species has reached a certain threshold concentration it will no longer be amplified exponentially, as a significant amount will have annealed to the double stranded form before the priming step. Thus, as each individual DNA species is amplified by the process to its threshold concentration, the rate of amplification of that species will start to tail off. Eventually, therefore, all DNA species will be present at the same concentration.

The length of the reannealing step will determine how much DNA is present at the threshold concentration. Preferably, therefore, the duration of the reannealing step will be determined empirically for each DNA population.

In the PCR normalisation process in general, the DNA primers may be adapted to prime selectively a sample of the total DNA population. By using primers which will only prime a sample of the population, only that sample will be amplified and normalised. The total quantity of DNA generated will thereby be reduced, which means that the cycling times can be kept low. This ensures that the method is applicable to complex DNA populations such as cDNA populations. In addition, a first primer can be used which is adapted selectively to prime a sample of the total cDNA population, and a second primer which is a general primer. Advantageously, the general primer is oligo dT (each primed cDNA will then be replicated in its entirely, as the oligo dT primer will anneal to the poly-A tail at the end of the cDNA).

In co-pending British Patent Application No 92 14873.3, filed by MRC 13th July, 1992, a new process is described which allows the study and identification of the individual members of a mixed or heterogenous population of nucleotide sequences perhaps of varying abundance. In preferred embodiments of the said process, the starting nucleic acid population is treated by:

(a) subjecting the nucleic acid to the action of a reagent,

preferably an endonuclease which has its cleavage and recognition sites separated, which reagent cleaves the nucleic acid so as to produce double stranded cleavage products the individual strands of which overlap at cleaved ends to leave a single strand extending to a known extent;

(b) ligating the cleavage products from (a) with a population of adaptor molecules to generate adaptored cleavage products, each of which adaptor molecules has a cleavage product end recognition sequence and the population thereof encompassing a range of adaptor molecules having recognition sequences complementary to a predetermined subset of the sequences of the cleavage-generated extending single strands; and

(c) selecting and separating only those adaptored cleavage products resulting from (b) which carry an adaptor of predetermined recognition sequence.

A preferred endonuclease for use in step (a) of the above process is Fok 1.

An important feature of this process is the use of adaptor molecules. The adaptors used must have "overhanging" fragment recognition sequences which reflect or are complementary to the extending cleavage-derived sequences which the adaptors are designed to react with. It is also preferred that the adaptors used should end with a 5' hydroxyl group. The avoidance of a 5' phosphate group removes the risk of inappropriate ligation involving the adaptors.

Adaptor molecules may also contain a portion permitting specific sequence selection and separation (as in step (c) of the process) when a sequence is attached to the adaptor. For example, an adaptor can carry biotin, thereby permitting advantage to be taken of the biotin/avidin
reaction in selecting and separating desired adaptored molecules. Additionally, adaptors preferably comprise a known and selected sequence such that specifically isolated adaptored molecules can be amplified by known techniques (such as PCR) using a primer complementary to the core sequence.

Preferably the adaptors are short double-stranded oligonucleotides which can be joined to the ends of cleavage products. They will have been chemically synthesised so that their sequence can be predetermined

and so that large concentrations can be easily produced. They may also be chemically modified in a way which allows them to be easily purified during the process. As mentioned above, ideally their 5' ends will be unphosphorylated so that once joined to fragments the adaptored end of the latter will no longer be able to participate in further ligation reactions.

It is preferred that the adaptor cleavage product end recognition sequences are on the 5' end of the longest oligonucleotide strand making up the preferred adaptor molecules, are at least 3 nucleotides in length and with totally random bases at the single-stranded position(s) two nucleotides in from the 5' end. This then allows selection to be performed both during the joining reaction and during subsequent priming reactions. Then, because the final degree of selection is a result of the product of the degrees of selection achieved at these two stages, maximum selection can be achieved per adaptor/primer available.

Adaptor strand extensions on the 5' end of the longest oligonucleotide also facilitate the use of modified oligonucleotides for separation purposes. Preferably, the short oligonucleotide will be modified at its 5' end. This has the double benefit of requiring just one modified oligonucleotide for all possible single-stranded extensions that are used, and also placing the modification at a position where it cannot interfere with ligation or subsequent priming reactions.

Although only one type of adaptor is required per ligation reaction, it is preferred that adaptors covering all possible reactions in a chosen subset of sequences be present, because then the opportunity for fragments in the chosen subset to ligate to each other is minimised. It is also preferred that the chosen specific adaptor, carrying a predetermined recognition sequence, should not only be different from the other adaptors in its single-stranded extension, but also different in the rest of its sequence since this allows orientation to be introduced which is useful in subsequent steps. It is therefore also preferred that this adaptor has a modified oligonucleotide to facilitate its separation with the cleaving products to which it joins.

The above "adaptoring" process can be used to generate categories or subsets of sequences by making some of the adaptors specific in some way, and selecting and separating as in step (c). In this way subsets of sequences can be provided depending upon the specific adaptor

chosen, e.g. for use in subsequent nucleotide sequencing. This facilitates, for example, the identification of a large population of sequences by permitting a rational approach to splitting such populations into subsets, each of which subsets can be examined in turn.

In the light of these developments, the present invention now provides a nucleic acid fragment encoding a gene product or portion thereof and comprising any one of:-

(a) a sequence selected from SEQ ID Nos 1 to 1193;

(b) an allelic variation of a sequence as defined in (a); or

(c) a sequence complementary to (a) or (b).

In another aspect, the invention provides a nucleic acid sequence as set out in any one of SEQ ID Nos 1 to 1193, or a complement or allelic variation thereof. Preferred sequences exhibit no more than 90% homology to a human sequence known per se.

In a further aspect, the invention provides a nucleic acid fragment comprising a portion of a sequence as defined above of sufficient size such that a probe of the same size and exhibiting complementarity to said portion can hybridise to said sequence. Preferably, such portions are at least 15 bases in length. It will be appreciated that minor mismatches in the aforesaid "complementarity" are not excluded provided hybridisation can still occur. In general, hybridisation conditions are within the choice of the skilled person, but reference can be made, for example, to the following: Melting temperature of hybrids - Bolton, E. T. and McCarthy, B. J. Proc. Natl. Acad. Sci, 48 p1390 (1962). Effect of formamide on lowering melting temperature - Casey, J. and Davidson, N., Nucleic Acids Res. 4, p1539 (1977). Effect of imperfect homology - Bonner, T. I. et al., J. Mol. Biol. 81, p123 (1973). General - Meinkoth, J. and Wahl, G. Anal. Biochem. 138, p267 (1984). Oligo hybridization and washing - Lathe, R. J. Mol. Biol. 183, P1 (1985).

The present invention also envisages DNA constructs comprising fragments or sequences as referred to above with a control or regulatory sequence.

The invention includes such DNA constructs using a gene system known in the art ligated to a sequence or fragment of the invention so as to enable, upon expression, the provision of a fusion polypeptide. Preferably, an endopeptidase recognition site is provided such that when the sequence or fragment is expressed it is expressed in frame with a known protein with the boundary being a cleavage site for an endopeptidase with a rare cutting site. The known protein can then be affinity purified, and the peptide corresponding to the fragment or sequence in accordance with the invention may be released by the endopeptidase. Alternatively, the whole protein can be used to raise antibodies which can then be screened for those directed at polypeptide corresponding to the fragment or sequence of the invention.

Since the present fragments and sequences can be used to produce, inter alia, corresponding genes, whether by isolating them, by synthesis or otherwise, such use and the resulting DNA fragments comprising genes are further aspects of the invention.

Yet another aspect of the invention is an expression vector comprising a fragment, sequence, gene-comprising DNA fragment, or DNA construct, as above, positioned such that that nucleic acid sequence which encodes the polypeptide corresponding to said fragment, sequence or DNA fragment is in operable reading frame with a control or regulatory sequence.

Other aspects of the invention are host cells incorporating a sequence, or fragment, or gene-comprising DNA fragment, or DNA construct, as above, as a heterologous part of the expressible genetic information of the cell. The production of such modified host cells can be achieved using methods known in the art. Such modified host cells can be used to express corresponding proteins, and these materials lend themselves in turn to the preparation of corresponding monoclonal or polyclonal antibodies using standard techniques.

Also included in this invention are such antibodies. Reference can be made, inter alia, to the following literature: Monoclonal antibodies, Cambell, A. M. Laboratory Techniques in Biochemistry, and Molecular Biology Ed. Burdon, R. H. and van Knippenberg, P. H. vol 13. Elsevier Amsterdam 1984. Goding, J. W., Monoclonal antibodies: Principles and Practice, 2nd Edition, academic Press, London 1986. Kipps, T. J. and Herzenberg, L. A., Handbook of Experimental Immunology : Applications of immunological methods in biomedical sciences, 4th edition Ed. Weir,

D. M. et al., p108 Blackwell scientific Publications, Oxford. Harlow, E and Lane, D. Antibodies, A Laboratory Manual, Cold spring harbor Laboratory, Cold Spring, New York.

Expression in an appropriate higher eucaryotic host may be important to ensure correct protein folding and also activity. Expression to avoid copurification of toxic products can sometimes be better performed in organisms approved for human consumption, eg prokaryotic Bacillus subtilis, eurkaryotic yeast, mammalian cows milk vectors, and other methods known in the art.

The invention also includes novel gene products or portions thereof encoded by a fragment, sequence or gene-comprising DNA fragment of the invention.

It will be appreciated that the sequences of the present invention collectively have utility based, inter alia, upon their common origin, and hence they can effectively be considered together rather than as separate entities. It is convenient to represent them as separate sequences, because this is how they were produced and serves as "punctuation" between the different functional entities which each sequence represents. However, the sequences could just as easily have been presented as a continuous sequence derived by placing them end to end in the order in which they were produced, with a separate indication of where the beginnings and ends of the component sequences are.

In contrast to investigations hitherto, where gene fragments (sequence fragments) could only be identified through some known characteristic [for example: their homology to a fragment which largely encodes amino acids identified by sequencing a previously isolated peptide or is the antisense of that coding sequence; or them having at least partial homology to previously characterised nucleic acids; or them having ability to encode expressed proteins which could later be detected by functional assays of the cells expressing those proteins or by using antibodies which had been previously raised against the proteins to detect their expression, Sambrook J., et al., Molecular Cloning CSH Press 1989], the sequences and fragments described by the present invention are entirely underivable and unpredictable from the prior art, but are nonetheless clearly of great value for various purposes.

Thus, such sequences, by comparing them to sequence databases, can be

used as a means for determining the existence of new members of existing gene families, new human genes when previously only non-human genes were known and new genes when previously no genes were known (Karlin, S. and Altschul, S. F. Proc. Natl. Acad. Sci. 87 p2264-2268 (1980)). In all cases, this allows the isolation of the corresponding genes and their products, and hence enables the manufacture of molecules of potential biological interest by recombinant means. Screening libraries of known materials or hitherto unexplored source materials for biological efficacy is now an important industrial activity in the search for new therapies and therapeutics. When new sequences have already been found to have counterparts in gene families or in non-human genes then knowledge about biological efficacy may already be apparant. For example, new receptors or receptor agonists/antagonists may exhibit differences to known instances of these molecules, and such differences could make them more suitable as therapeutics by, for example, exhibiting binding characteristics which are more in keeping with avoidance of toxicity. Reference can be made, for example, to polymorphic dopamine receptors and the implications for mental health (Iversen, L. Nature 358, p109 (1992), and Van Tol, H. M. M. et al., Nature 358, p149-152 (1992)). Where absolutely required, realisation of full length cDNAs for expression can be achieved by using the sequences to screen (by hybridisation) suitable cDNA libraries containing full length clones (D'Alession, J. M., et al., Focus (Gibco B.R.L) 9 p1 (1987)). Alternatively, the sequences can be used to design primers suitable for obtaining the missing sequences by PCR or other amplification methods (Frohman, M. A., Dush, M. K. and Martin, G. R., Proc. Natl. Acad. Sci. 85 p8998-9002 (1988)).

Appropriate use of the sequence fragments in antisense or triple helix (Griffin et al., Science 245 p967-971 (1989)) applications will be useful for identifying manipulable targets related to disease. For example, viruses have been inhibited by antisense RNA to their mRNAs (Chang, L-J., and Stoltfuz, C. M. J. Virol. p921-974 (1987)). A similar effect could be achieved by targetting the expression of cellular proteins which are essential for growth or maintenance of the virus.

Partial or full length cDNAs have great utility once expressed. The manner of expression can be selected by one skilled in the art to suit the intended application. Expression of full length cDNAs is typically required for biological activity. Procaryotic, and lower or higher eucaryotic hosts may be selected as the host for expression and higher

eucaryotes may be preferred to ensure correct modifications, for example, glycosylation in vivo, when this proves to be important. Expression can be ensured by situating the cDNA appropriately to signals for expression (Amann, E. and Brosius, J. Gene 40 p183 1985), Shimuzu, Y et al., Gene 65, p141 (1988), Straus, D. and Gilbert, W. Proc. Natl. Acad. Sci. 82, p2014 (1985)). Such signals may include a promoter for transcription, which may itself be regulatable.

The proteins thus-expressed can be screened for activities of therapeutic or commercial value. It may be that the proteins have to be first isolated for this purpose or can be assayed in situ. It may be desirable that some means of stabilising the expressed protein is employed. This can be achieved, for example, (and as indicated earlier) by expressing in frame as part of a fusion polypeptide (Smith, D. B., et al., Proc. Natl. Acad. Sci. 83 p8073 (1986)).

Useful antibodies can be raised against the expressed proteins. It is commonly not an absolute requirement that full length proteins are produced, although this may influence the quality of the antibodies produced. Peptides as short as 8 or 9 amino-acids in length can be used as antigens (Germain R., N. Nature 353 pp605-607 (1991), Rudensky, A., Y., et al., Nature 353 p622-627 (1991)). Immunogenic peptides could simply be synthesised using the amino-acid sequence translated from a sequence or fragment of this invention. It is desirable, although not absolutely required, that some means of producing purified antibodies is adopted. When fusion polypeptides are used to raise antibodies, an affinity matrix specific for the generic part of the protein allows the fusion polypeptide to be immobilised (Smith, D. B., et al., Proc. Natl. Acad. Sci. 83 p8073 (1986)). The immobilised polypeptide can then be used to affinity purify the antibodies. Antibodies to both the generic part of the fusion polypeptide and the part of interest are produced. When these need to be discriminated between, a different affinity column can be used to remove only those antibodies specific for the generic part of the polypeptide. Alternatively, and as mentioned earlier, it can be arranged that the boundary between the two separate protein components of the fusion polypeptide has the recognition sequence for an endopeptidase with a rare cutting site. The peptide of interest can then be released from the affinity purified polypeptide by the action of the endopeptidase (Nagai, K., and Thogersen, H., C. Methods Enzmol. 153 p461-481 (1987). Another alternative is raise monoclonal antibodies against the purified protein.

The antibodies can be used for localising in situ, or quantifying in samples through, for example, ELISA or RIA assays, peptides against which they were raised. These uses are particularly beneficial when the results of the assays can be correlated to a disease condition, eg cancer. For example tumour markers may be found and used to target therapeutic agents. The antibodies can also be used to detect or monitor markers of undifferentiated growth, infection, cardiovascular or immune disease or a therapeutic response. When the antibodies recognise cell surface proteins they can be used in isolation or in combination to isolate particular populations of cells. These in turn can be used to isolate yet more cDNAs which will be enriched for yet more of such surface markers for the population, which, if similarly screened, will permit yet further subdivision of the population. Ultimately, panels of antibodies which can describe particular disease states will accrue. Such antibodies could be tailored for forensic applications as well as diagnostic purposes and disease monitoring.

The sequences or fragments can also be used for genetic analysis and mapping, for example, to diagnose the likelihood that a given individual is predisposed towards a given genetic disease. In the event of a sequence co-locating, genetically, with a disease gene, it can be used for the derivation of new disease therapies bases upon precise genetic knowledge. Such therapies can include, for example, the techniques of so-called "gene therapy" (Dusty Miller, A. Nature 357 p455-460 (1992)).

Antibodies can be produced against the protein of a genetic disease with sufficient discriminating power to discriminate between diseased and non-diseased states (Caskey, T. Genome Sequencing Conference, Hilton Head, S. Carolina (1991)). This would be useful for reducing the dependence of such tests on nucleic acid-based screens. Such antibodies also have the advantage of allowing detection of faulty expression of the protein, for example levels of expression which may be important for development of the disease in slow onset conditions.

Also very important is that not all cDNAs are likely to be found by conventional means, whereas the present sequences are, in one sense, "comprehensive". The use of the class of cDNAs which corresponds of necessity to truncated clones increases the chances that part of a cDNA will be cloned free of any sequences that could otherwise compromise it from being cloned. Sequence obtained can then be used to generate PCR primers from which the remainder can be obtained without having to

clone.

This invention will now be further described and illustrated by means of the following Examples.

All oligonucleotides used in these Examples were synthesised Trityl on using an ABI 380B DNA Synthesizer according to the manufacturers instructions. Purification was by reverse phase HPLC (see, for example, Becker, C., R., et al., J. Chromatography 326, p293-299 (1985)).

Example 1

Human brain and adrenal tissues were obtained from a mixture of 12 to 15 week menstrual age foetuses and then snap frozen in liquid nitrogen before storing in bijou bottles in a -80°C freezer. The two types of tissue were used separately, directly from the freezer, to prepare cDNA from which restriction fragments were generated for sorting into subsets. 1g portions of each of the separate tissues were homogenised, using an Ultra-Turrax T25 Disperser (Janke and Kunkel, IKA-Labortechnik), on ice in the presence of 4M guanidinium isothiocyanate to solubilise macromolecules. RNA was isolated from each homogenate by using centrifugation to sediment it through caesium trifluoroacetate. This was performed using the Pharmacia kit according to the manufacturer's instructions, except that centrifugation was performed for 36 hours and the RNA obtained was finally desalted and concentrated by performing two ethanol precipitations in succession with two 70% ethanol washes after each precipitation. In each case, polyA$^+$ (mRNA) was isolated from 200 to 400 μg of the total RNA by binding it to magnetic oligo-dT coated beads (Dynal). Solution containing unbound material was removed from the beads, which were washed, and then mRNA eluted directly for use. mRNA isolation was performed in accordance with the manufacturer's instructions. Yields of RNA from the beads were between 1 and 3% of the total RNA. 2 to 4 μg of the eluted RNA were used for cDNA synthesis. cDNA synthesis was performed according to the method of Gubler, U and Hoffman, (B. J. Gene 25 p263 (1983) using a Pharmacia kit according to the manufacturer's instructions. OligodT was used to prime the first strand cDNA synthesis reaction. The cDNA was purified by extracting twice with phenol/chloroform and then low molecular weight solutes including nucleic acids below ca. 300 bases were removed by passing the cDNA reaction mixture through a Pharmacia S400 spun column used according to the manufacturer's